Replace this file with `prentcsmacro.sty` for your meeting, or with `entcsmacro.sty` for your meeting. Both can be found at the ENTCS Macro Home Page.

Bio-PEPA for epidemiological models

Federica Ciocchetta[a,1], Jane Hillston[b,2]

[a] *The Microsoft Research - University of Trento Centre for Computational and Systems Biology, Trento, Italy*

[b] *Laboratory for Foundations of Computer Science, The University of Edinburgh, EH8 9AB Edinburgh, Scotland*

Abstract

Many models have been defined in order to describe the evolution of a disease in a population. The modelling of diseases is helpful to understand the mechanisms for their spread and to predict their future evolution. Most of the models in the literature are defined in terms of systems of differential equations and only a few of them propose stochastic simulation for the analysis.
The main aim of this work is to apply the process algebra Bio-PEPA for the modelling and analysis of epidemiological models. As Bio-PEPA has been originally defined for biochemical networks, we define a variant of it suitable for representing epidemiological models. Some features of Bio-PEPA are useful in the context of epidemiology as well: location can abstract spatial structure and event can describe the introduction of prophylaxis in a population infected by a disease at a given day. Concerning the analysis, we can take advantage of the various kinds of analysis supported by Bio-PEPA, such as, for instance, stochastic simulation, model checking and ODE-based analyses. In particular, the modeller can select the most appropriate approach for the study of the model and analysis techniques can be used together for a better understanding of the behaviour of the system.
In this paper we apply Bio-PEPA to the study of epidemiological models of avian influenza, based on different assumptions about the spatial structure and the possible kind of treatment. These models demonstrate that Bio-PEPA has several features that facilitate epidemiological modelling.

Keywords: Process algebra, epidemiological models, modelling, stochastic simulation, model checking, ordinary differential equations

1 Introduction

The outbreak and spread of epidemics have been studied for many years and many models have been defined in order to describe the evolution of a population infected by one or more diseases [1,2,25,6,23,8,18]. The modelling of diseases is helpful to understand the mechanisms for their spread, to predict the future course of an outbreak and to evaluate strategies to control it. The possibility to predict the evolution of the disease can help scientists to evaluate plans to handle it and may have a significant impact on the extinction (i.e. eradication) of a particular epidemic.

Most epidemiological models have been defined in terms of ordinary differential equations (ODEs) or partial differential equations (PDEs) and just a few stochastic models have been proposed. Differential equations are generally a good approach when the population can be assumed to be large and the focus is on the average behaviour. However, when the population is small and characterized by a certain variability, stochastic models can be a more faithful representation of the system and are therefore more appropriate.

[1] Email: ciocchetta@cosbi.eu

[2] Email: jeh@inf.ed.ac.uk

Process algebras were originally defined in the context of concurrent systems in computer science. They offer a formal model of a system in terms of interacting components, and support a compositional approach to model construction. From the perspective of epidemiological modelling they offer the possibility of describing the individuals in the population as distinct components with precise specifications of the interactions between them. Recently, there have been several explorations of process algebra modelling in the epidemiological context [33,28,27,4,7]. Most of these have focussed on the derivation of ODE-based models from the process algebra description, and stochastic simulation is considered only in [4]. Furthermore, simple epidemiological models are presented in these papers.

In recent years there has been considerable interest in applying process algebra modelling to biochemical networks [30,17,32,29,10]. In several cases process algebras have been developed specifically for this purpose, using slightly different assumptions about how components interact than would be standard in the computing context. In this work we apply one such process algebra, Bio-PEPA [15,16]. This is a modification of the stochastic process algebra PEPA which was originally developed for performance modelling of computer and communication systems [24]. The new language is equipped with functional rates to allow the rate of an action to depend on the current state of the system, classification of actions according to their impact on the number of entities and locations for segregating actions.

Since Bio-PEPA has been originally proposed for biochemical networks, we need to define a variant of it that is more suitable for describing epidemiological models. In particular, some details about biochemical species and locations are eliminated. Note that several features of Bio-PEPA remain useful in this new context. Indeed functional rates can express complex dynamic laws and locations can abstract spatial structures. Furthermore, Bio-PEPA has been enriched with events [11], which can represent, for instance, the introduction of prophylaxis in a population infected by a disease at a given day.

Concerning the analysis, from a Bio-PEPA model we can derive an associated system of ODEs, a model for stochastic simulation and a PRISM model [26,31] for model checking. Access to a variety of analysis techniques can help develop a better understanding of the behaviour of the system and allows the modeller to select the most appropriate approach for the study of the model considered.

We apply our variant of Bio-PEPA to modelling and analysis of epidemiological models for avian influenza [19], based on different assumptions about the spatial structure and the kind of treatment given (i.e. prophylaxis or introduction of a drug). Specifically, we start by considering a simple model describing the disease without locations and without treatment, then we extend it with locations and finally we add treatment and the possibility of drug resistance. In the study of these models the following quantities are of interest [19]: the maximum number of infective individuals with respect to time (*the epidemic's peak value*), the time when it occurs (*the epidemic's peak time*), *total number of instantaneous infectives, cumulative number of infectives generated during epidemic* and *the total number of resistant infectives*. Furthermore, we would like to investigate the impact of different population structures on the outcome of influenza and the effect of treatment and prophylaxis on drug resistance. In the simplest models the stochastic and deterministic analyses are in full agreement. However, when spatial structure is taken into account the variability of the system increases and the use of stochastic simulation seems more appropriate and essential to capture the behaviour of the system.

The rest of the paper is structured as follows. An introduction to epidemiological models and a description of Bio-PEPA for biochemical networks are reported in Sect. 2 and Sect. 3, respectively. The variant of Bio-PEPA for epidemiological models is introduced in Sect. 4. In Sect. 5 the Bio-PEPA systems describing three models concerning the avian influenza are presented and some analyses are reported. Finally, in Sect. 6 we report some concluding remarks.

2 Epidemiological models

There is a vast literature of models describing populations infected by one or more diseases [1,2,25,6,23,8,18]. These models differ from each other in terms of the dynamics in the absence of disease (i.e. the total population can be constant or birth and death rates can be prescribed by some laws) and in terms of assumptions about the infection. For instance, the disease can be transmitted directly by contact with an infective individual or through a vector, and it can be assumed that offsprings of infected parents can be born infected or not. Furthermore, there is a distinction in terms of the number of diseases active in a population and the relationship between them.

With respect to a disease, the population can be divided into classes: *susceptibles (S)*, those who are *infective (I)*[3] and those who have *recovered* and are immune (*R*). In addition to these, we can have other classes, describing, for instance, symptomatic or asymptomatic infectives, treated, untreated or immune individuals. According to the kind of disease and the classes in which the population is divided, the following classification of models has been proposed: *SI*, in which susceptibles can be infected but do not recover, *SIS*, in which infectives can recover and become susceptible again, *SIR*, in which infectives can recover and remain immune to the disease and *SIRS*, where infectives can recover and recovereds can become susceptible again. More complex models are obtained by considering other classes and assumptions.

In epidemiology (as in all ecological systems), spatial structures can have a large impact on the evolution of a population and on the outcome of a disease [19,22]. Generally, an abstract view of space is sufficient to describe the spatial evolution of the epidemic. The term *metapopulation* is used to indicate a population distributed over a number of *patches* or *subpopulations*, i.e. groups of hosts in the model. Individuals can migrate from one patch to another and this can be described by a *migration matrix*, that determines the topology and the strength of the connections between the patches. The dynamics of a metapopulation are a function of both within-population dynamics and among-population dynamics.

Different population structures, depending on the number of subpopulations and how they are connected, can be defined. Some possible structures, reported in Fig. 1, are:

- *island-type*: all patches are equally accessible from all other patches in a single step;
- *spider-type*: individuals can travel between the central patch and outer patches in a single step. All the possible movements are via the central patch;
- *necklace-type:* individuals can move only to adjacent patches in one step;
- *loop-type*: individuals can move only to adjacent patches in one step and all patches are connected to another two patches in a loop.

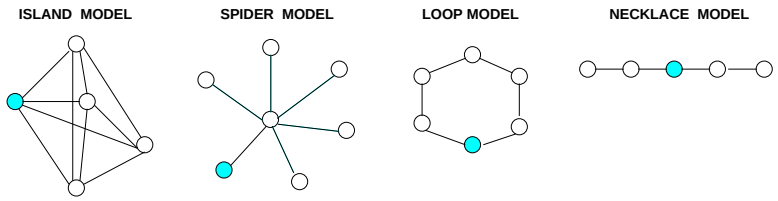

Fig. 1. Example population structures. The blue circle illustrates the patch with disease initially in the experiments in Section 5.

An example of a population divided into subpopulations with migration between them is illustrated in Fig. 2. The population (blue circle) consists of 5 subpopulations (white circles). Each subpopulation is

[3] In this work we assume that infected individuals correspond to infective ones.

composed of different numbers of susceptibles, infectives and recovereds. The arrows between the circles represent the possible migration of individuals from one patch to another.

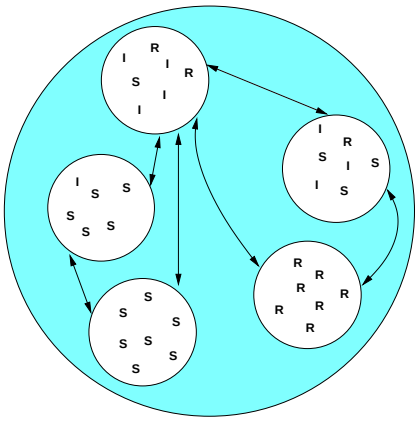

Fig. 2. Example of subpopulations. The whole population (blue circle) is divided into some subpopulations (white circles), arrows represent migration of individuals between subpopulations. S, I, R stand for susceptible, infective and recovered individuals.

3 Bio-PEPA

In this section we give a short description of Bio-PEPA [15,16], a language that has recently been developed for the modelling and analysis of biochemical systems. The main components of a Bio-PEPA system are the *species components*, describing the behaviour of each species, and the *model component*, describing the interactions between the various species. The species initial amounts are given in the model component.

The syntax of the Bio-PEPA components is defined as:

$$S ::= (\alpha, \kappa)\ \mathsf{op}\ S \mid S + S \mid C \quad \text{with } \mathsf{op} = \downarrow \mid \uparrow \mid \oplus \mid \ominus \mid \odot \qquad P ::= P \underset{\mathcal{L}}{\bowtie} P \mid S(x)$$

where S is the *species component* and P is the *model component*. In the prefix term $(\alpha, \kappa)\ \mathsf{op}\ S$, κ is the *stoichiometry coefficient* of species S in reaction α, and the *prefix combinator* "op" represents the role of S in the reaction. Specifically, $\downarrow$ indicates a *reactant*, $\uparrow$ a *product*, $\oplus$ an *activator*, $\ominus$ an *inhibitor* and $\odot$ a generic *modifier*. We can use "α op " and "(α, κ) op " as abbreviations for "$(\alpha, 1)$ op S" and "(α, κ) op S", respectively. The operator "+" expresses the choice between possible actions, and the constant C is defined by an equation $C \overset{def}{=} S$. The process $P \underset{\mathcal{L}}{\bowtie} Q$ denotes synchronisation between components P and Q, the set $\mathcal{L}$ determines those activities on which the operands are forced to synchronise, with $\underset{*}{\bowtie}$ denoting a synchronisation on all common action types. In the model component $S(x)$, the parameter $x \in \mathbb{R}$ represents the initial amount of the species.

Recently Bio-PEPA has been extended to incorporate events [11] and biological locations [14].

Events are constructs that represent changes in the system due to some triggering conditions. This allows biochemical perturbations to the system to be represented, such as the timed introduction of reagents or the modulation of system components by external stimuli. A Bio-PEPA event has the form (*id, trigger, event_assignment, delay)*, where *id* is the event name, *trigger* is a mathematical expression involving the components of the Bio-PEPA model and/or time, *event_assignment* is a list of assignments causing some changes to elements in the system, and *delay* is either 0 (*immediate events*) or a positive real value (*delayed events*). Events are added to the language as a set of elements and the rest of the syntax is unchanged in

order to keep the specification of the model as simple as possible. Furthermore, this approach is useful when the same biochemical system is studied under different experimental regimes as the list of events can be modified without any changes to the rest of the system.

Locations represent both biological compartments and membranes. They form a static hierarchy (i.e. they have a fixed structure) but they can change size with time. The notation $C@L$ indicates that the species represented by the component C is in the location L. If we have just one location, it is treated implicitly and we simply write C. The locations are defined as follows.

Definition 3.1 Each location is described by "L : s *unit*, *kind*", where L is the (unique) location name, "s" expresses the size and can be either a positive real number or a more complex mathematical expression depending on time t; the (optional) "*unit*" denotes the unit of measure associated with the location size, and "*kind*" $\in \{\mathbf{M}, \mathbf{C}\}$ expresses if it is a membrane or a compartment, respectively.

A Bio-PEPA *system* representing a biochemical network consists of a set of sequential components, a model component, and a context (kinetics rates, parameters, locations, events, etc.).

Definition 3.2 A Bio-PEPA system $\mathcal{P}$ is a 7-tuple $\langle \mathcal{L}, \mathcal{N}, \mathcal{K}, \mathcal{F}_R, Comp, P, Events \rangle$, where: $\mathcal{L}$ is the set of locations, $\mathcal{N}$ is the set of auxiliary information, $\mathcal{K}$ is the set of parameters, $\mathcal{F}_R$ is the set of functional rates, $Comp$ is the set of species components, P is the model component and $Events$ is the set of events.

Bio-PEPA is given an operational semantics [15,16]. In this context species amounts are abstracted by discrete levels, representing intervals of values. There are two relations over the processes: the *capability relation*, which supports the derivation of qualitative information, and the *stochastic relation*, defined in terms of the capability relation and equipped with rates for the associated action types.

The capability relation is $\rightarrow_c \subseteq C \times \Theta \times C$, where C is the set of model components and Θ is the set of labels. The labels $\theta \in \Theta$ are defined as $\theta := (\alpha, w)$, where w is a list recording the species that participate in the reaction and is defined as $[S : op(l, \kappa, species\_location)] \mid w :: w$, with S the name of the species component, l the level, κ the stoichiometry coefficient and $species\_location$ is the location of species. The relation $\rightarrow_c$ is the minimum relation satisfying the rules reported in Table 1.

The stochastic relation is $\overset{\gamma}{\mapsto} \subseteq \tilde{\mathcal{P}} \times \Gamma \times \tilde{\mathcal{P}}$, where the label $\gamma \in \Gamma$ is defined as $\gamma := (\alpha, r_\alpha, reaction\_location)$, with $r_\alpha \in \mathbb{R}^+$ and $\tilde{\mathcal{P}}$ is the set of well-defined Bio-PEPA systems [15]. In this definition r_α is the rate associated with the action type α and is calculated using the functional rate and the information in w. This rate represents the parameter of a negative exponential distribution. The dynamic behaviour of processes is determined by a *race condition*: all activities enabled attempt to proceed but only the fastest succeeds. The element *reaction_location* indicates the location of the reaction. For instance, it can be a compartment, if all the reagents are in a given location, or it can be of the form $L_1 \Rightarrow L_2$, if the reaction is a transport of a species from L_1 to L_2. The relation $\overset{\gamma}{\mapsto}$ is defined as the minimal relation satisfying the rule

$$\texttt{Final} \quad \frac{P \xrightarrow{(\alpha_j, w)}_c P'}{\langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, Comp, P \rangle \xmapsto{(\alpha_j, r_\alpha, reaction\_location)} \langle \mathcal{V}, \mathcal{N}, \mathcal{K}, \mathcal{F}, Comp, P' \rangle}$$

The Bio-PEPA language is supported by a suite of software tools which automatically process Bio-PEPA models and generate other representations in forms suitable for different kinds of analysis [16,13,5]. These tools capture mappings from Bio-PEPA to differential equations, stochastic simulation models [21],

`prefixReac`	$((\alpha,\kappa,)\downarrow S)(l) \xrightarrow{(\alpha,[S:\downarrow(l,\kappa,species\_location)])}_c S(l-\kappa) \quad \kappa \le l \le N$
`prefixProd`	$((\alpha,\kappa)\uparrow S)(l) \xrightarrow{(\alpha,[S:\uparrow(l,\kappa,species\_location)])}_c S(l+\kappa) \quad 0 \le l \le (N-\kappa)$
`prefixMod`	$((\alpha,\kappa)\, op\, S)(l) \xrightarrow{(\alpha,[S:op(l,\kappa,species\_location)])}_c S(l)$ with $op = \odot, \oplus, \ominus$ and $0 < l \le N$ if $op = \oplus$, $0 \le l \le N$ otherwise

$$\texttt{choice1}\quad \frac{S_1(l) \xrightarrow{(\alpha,w)}_c S_1'(l')}{(S_1+S_2)(l) \xrightarrow{(\alpha,w)}_c S_1'(l')} \qquad \texttt{choice2}\quad \frac{S_2(l) \xrightarrow{(\alpha,w)}_c S_2'(l')}{(S_1+S_2)(l) \xrightarrow{(\alpha,w)}_c S_2'(l')}$$

$$\texttt{constant}\quad \frac{S(l) \xrightarrow{(\alpha,S:[op(l,\kappa,species\_location)])}_c S'(l')}{C(l) \xrightarrow{(\alpha,C:[op(l,\kappa,species\_location)])}_c S'(l')} \text{ with } C \stackrel{def}{=} S$$

$$\texttt{coop1}\quad \frac{P_1 \xrightarrow{(\alpha,w)}_c P_1'}{P_1 \underset{\mathcal{L}}{\bowtie} P_2 \xrightarrow{(\alpha,w)}_c P_1' \underset{\mathcal{L}}{\bowtie} P_2} \text{ with } \alpha \notin \mathcal{L} \qquad \texttt{coop2}\quad \frac{P_2 \xrightarrow{(\alpha,w)}_c P_2'}{P_1 \underset{\mathcal{L}}{\bowtie} P_2 \xrightarrow{(\alpha,w)}_c P_1 \underset{\mathcal{L}}{\bowtie} P_2'} \text{ with } \alpha \notin \mathcal{L}$$

$$\texttt{coop3}\quad \frac{P_1 \xrightarrow{(\alpha,w_1)}_c P_1' \quad P_2 \xrightarrow{(\alpha,w_2)}_c P_2'}{P_1 \underset{\mathcal{L}}{\bowtie} P_2 \xrightarrow{(\alpha,w_1::w_2)}_c P_1' \underset{\mathcal{L}}{\bowtie} P_2'} \text{ with } \alpha \in \mathcal{L}$$

Table 1
Axioms and rules for Bio-PEPA.

continuous time Markov chains (CTMC) with levels [12] and PRISM models [26] [4] .

4 A variant of Bio-PEPA for epidemiological models

As we extend the use of Bio-PEPA from biochemical networks to epidemiological models there are some key issues which need to be considered:

- In epidemiology each species will correspond to a set of individuals exhibiting the same behaviour (e.g. susceptible, infective or recovered individuals). In this new context the role of a "species" with respect to an action does not have the same significance as in the biochemistry. Nevertheless, we can still use the operators for the species role to indicate that the species decreases, remains invariant or increases in an interaction. The distinction between inhibitors, enzymes and generic modifiers is no longer needed and it is sufficient to retain only the generic modifier.
- It is possible to have an interaction of the form $I+S \to 2I$ or similar, where a species (I) is present on both sides of the interaction with different stoichiometry or multiplicity. Note that this cannot be represented in Bio-PEPA [5] .
- *Spatial structures* are often present in epidemiological models. These can be translated as Bio-PEPA locations; however the distinction between membranes and compartments is not meaningful here.

[4] At the moment, in the case of models with events, just stochastic simulation and ODE numeric integration are supported.

[5] In the biochemical context if a species is present on both side of a reaction it must be a modifier and the two stoichiometric coefficients must be the same.

4.1 The syntax of Bio-PEPA for epidemiological models

A Bio-PEPA model for epidemiological system is described by the following syntax:

$$S ::= (\alpha,\kappa) \downarrow S \mid (\alpha,\kappa) \uparrow S \mid (\alpha,(\kappa_1,\kappa_2)) \odot S \quad \mid S + S \mid C \mid S @ L \qquad P ::= P \underset{I}{\bowtie} P \mid S(x)$$

where the terms have the same meaning as explained in Sec. 3 [16,14]. With respect to the standard Bio-PEPA syntax, we do not have the operators for the enzyme and the inhibitor and we use a new prefix $(\alpha,(\kappa_1,\kappa_2)) \odot$ for species present on both sides of an interaction. The pair (κ_1,κ_2) represent the species' two multiplicity coefficients, before and after the interaction, respectively.

Note that the definition of location is as reported in the previous section, but here we do not have the attribute *kind* because it not meaningful in this context. Thus the one kind of location becomes implicit.

Definition 4.1 Each location is described by "$L : s\ unit$", where L is the location name, "s" expresses the size and can be either a positive real number or a more complex mathematical expression depending on time t; the (optional) "*unit*" denotes the unit of measure of the location size.

A Bio-PEPA system is defined as previously (i.e. Def. 3.2). The operational semantics for this variant of Bio-PEPA differs from the one reported in the previous section only in the axiom describing modifiers; specifically, the axiom `prefixMod` is replaced by the following:

$$\texttt{prefixGenMod}\ \ (\alpha,(\kappa_1,\kappa_2)) \odot S(l) \xrightarrow{(\alpha,[S:\odot(l,(\kappa_1,\kappa_2),species\_location)])}_c S(l-\kappa_1+\kappa_2)\,.$$

Note that in epidemiological models species are expressed in terms of number of individuals; this feature is reflected in the Bio-PEPA model too.

4.2 Abstraction

The translation of epidemiological models into Bio-PEPA is based on the following correspondences:

- Each *subpopulation/patch* is abstracted by a *location*.
- Each *species* is represented by a *species component*, whose subterms describe its interaction capabilities.
- Each *interaction* is represented by an *action type*. The dynamics are described by a *functional rate*.
- The model component represents how the species interact and contains information about the initial state.

4.3 Mappings to analysis

As in the context of biochemical networks, we can take advantage of the mappings which have been defined from Bio-PEPA to various mathematical models for analysis, including stochastic simulation [21], analysis based on ordinary differential equations (ODEs), numerical solution of the CTMC with levels [12] and stochastic model checking using PRISM [31,26]. It is worth noting that each of these analyses can aid understanding complementary aspects of the behaviour of the system. Furthermore, when two analyses overlap in scope, the results obtained can be used for verification. Specifically, stochastic simulation and numerical integration of ODEs are useful to study the temporal evolution of the system, whereas model checking can be applied to investigate some properties of the system that cannot be easily seen from simulation. In particular, stochastic simulation is appropriate when the system is composed of small numbers of species and

we are interested in capturing the variability within the system; on the other hand the ODE model is suitable when the number of entities is large and the focus is on the average behaviour of the system.

Below we report just some observations, the interested reader is refered to [16] for details. The mapping from a Bio-PEPA model to the corresponding systems of ODEs is the same as for the standard Bio-PEPA [16], except for a small change to take account of the new generic modifier with two multiplicity coefficients. The mapping to stochastic simulation is identical to the one proposed in [16], but the conversion from concentrations to number of individuals is not necessary since the model is already expressed as numbers of individuals. Finally, the mapping to CTMC with levels and to a PRISM model [31,26] are identical to those for standard Bio-PEPA [12] with the addition of the case of species present on both side of a interaction due to the new generic modifier. Species are represented in terms of levels, but here, differently from [12], each level abstracts an interval of integer values instead of an interval of concentrations, since the models are expressed in terms of number of individuals. Given a species amount, the corresponding level is derived by dividing the amount value by the step size H. The step size expresses the granularity of the system. The information about the step size is contained in the set $\mathcal{N}$. A maximum amount (M_{Max}) is assumed for all the species. This allows us to have a finite CTMC and therefore guarantees that the CTMC-based analysis is feasible.

In PRISM it is possible to specify quantitative properties of the system using the temporal logic *CSL* (Continuous Stochastic Logic) and rewards [3,31]. In the context of epidemiological systems some properties of interest are:

- *Probability of coexistence of the species at or before time t.*

$$\mathcal{P}_{=?}[\, true \;\; \mathrm{U}^{\leq t} \;\; S > 0 \;\; \& \;\; I > 0 \;\; \& \;\; R > 0]$$

- *Probability of extinction of the disease at or before time t*:

$$\mathcal{P}_{=?}[\, true \;\; \mathrm{U}^{\leq t} \;\; I = 0\,]$$

- *Long run probabilities (steady states) of coexistence of the species and extinction of infection*:

$$\mathcal{S}_{=?}[\, S > 0 \;\; \& \;\; I > 0 \;\; \& \;\; R > 0] \qquad \mathcal{S}_{=?}[I = 0]$$

- *Expected time until extinction of the disease.*

$$\mathcal{R}^{time}_{=?}[\mathrm{F} \;\; I = 0\,]$$

where the operator $\mathcal{R}_{=?}$ is used to express a reward and F indicates the case of a reachability reward.

In the properties above we assume three subclasses for the population (S, I, R), but the properties can be easily extended to models with more subclasses.

5 Examples: models of H5N1 avian influenza

In this section we apply Bio-PEPA to the study of models from [19] concerning the spread of *H5N1 avian influenza*. The Bio-PEPA Workbench [34] is used for the analysis. Specifically, we consider the mapping to the Dizzy simulator tool supported by the Bio-PEPA Workbench and use the version of Dizzy developed at the University of Edinburgh [20] for both numeric integration of ODEs and stochastic simulation. The ODE

solver *ODEtoJava-dopr54-adaptive* (with a variable time-step size that is controlled by an adaptive method involving a formula for estimating the error) is used for the numerical integration whereas Gillespie's direct method [21] is the algorithm chosen for stochastic simulation. Furthermore, we consider the mapping to the PRISM model checker [31] and we use it to verify some properties for one of the models.

Both stochastic simulation and numerical integration of the ODE models are fast: a few seconds for all the models (in the most complex case we have seven seconds for 100 stochastic simulation runs and less than one second for the ODE numerical integration). The verification of properties in PRISM is relatively time-consuming even for the simplest of the models considered (about ten minutes for each property).

The results obtained are compared with the known behaviour of the disease reported in [19]. In that paper the authors use systems of ODEs and bespoke simulation models.

5.1 Assumptions and definitions

The time unit in all our models is one day and this is reflected in the graphs. The models considered in this work are based on the following assumptions:

- The total population is constant and composed of $N = 500$ individuals. The assumption of a constant population is appropriate as we limit our attention to a short period of time (50 days).
- The basic model is an $SIIsR$ model: we have susceptibles (S) who can be infected by direct contact with infectives and infectives who can recover (R). The infectives can be of two types: I, *asymptomatic*, and Is, *symptomatic*. Both are infectious, but their infectiousness varies, and only those who develop clinical symptoms will be treated, when treatment is provided.
- The transmission of the influenza is by direct contact between an infective and a susceptible.
- We consider all the spatial arrangements discussed earlier — *island-*, *spider-*, *loop-* and *necklace-type* — but due to space limitations we present only island-type and necklace-type space. The population is divided into $n = 5$ *patches*. Each patch has (initially) N/n individuals. All the patches have the same dynamics and the migration rates are the same for all the connected patches.

From the basic model we can obtain other more complex models, adding the possibility of *treatment* and *resistance to the drug*. A *resistant virus* is a virus resistant to the drug and *resistant cases* are individuals with drug resistant viruses and thus the drug has no effect on them. Concerning treatment, the following two possibilities are considered:

- *Treatment for symptomatic infectives.* Infected people can receive drugs and become less infective and recover faster. Treatment for symptomatic infectives is a continuous intervention.
- *Prophylaxis.* Both asymptomatic infectives and susceptibles can receive prophylaxis. Susceptibles treated with prophylaxes become less susceptible to infection, while prophylaxed infectives are less infectious and recover faster. Prophylaxis is made at one (or more) specific time.

Treatment and prophylaxis do not work in the case of resistant viruses.

If we consider treatment and drug resistance, we have to define some specific subclasses for the population. Specifically, infectives can be divided into *treated (tr)* or *not*, with *resistance (r) or not*, *symptomatic (s)* or *not* and all the possible combinations of these cases. Susceptibles can receive *prophylaxis (pr) or not.* A schema of the species involved is reported in Fig. 3. The eight classes of infectives can be represented on a cube. The axes are *resistance, treatment, symptoms*. New infected individuals enter the cube from the bottom (no symptoms), as the symptoms do not appear immediately. On this side the new infectives go to

the resistant edge if they are infected by an individual with resistant viruses and they go to the treated edge, if they have received prophylaxis. The classes are: S (susceptibles), S_{pr} (susceptibles receiving prophylaxis), I (infectives, untreated and asymptomatic), I_s (infectives, untreated, symptomatic), I_r (infectives, untreated, with resistant virus, asymptomatic), I_{tr} (infectives, treated, asymptomatic), I_{sr} (infectives, untreated, with resistant virus, symptomatic), I_{str} (infectives, treated, symptomatic), I_{rtr} (infectives, treated, with resistant virus, asymptomatic), I_{srtr} (infectives, treated, with resistant virus, symptomatic), R, recovered.

The notation X_i is used to indicate the class X (from the classes listed above) in the patch i. In the analysis, we are interested in the total number of individuals of a class in all the possible locations. So for each class, the total number of individuals is $X = \sum_{i=1}^{n} X_i$, where n is the number of locations.

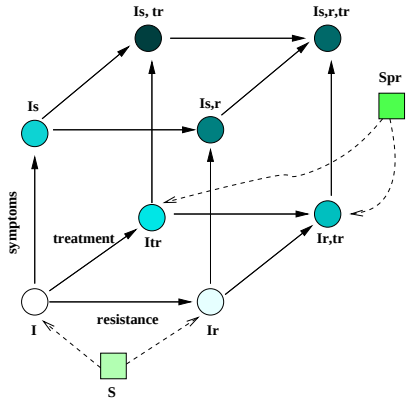

Fig. 3. Population reactions for the metapopulation model with treatment and drug resistance.

In this paper we consider the following kinds of models:

(i) *Simple SIIsR model with single (implicit) location*;

(ii) *Simple SIIsR models with locations*;

(iii) *SIIsR models with locations, extended with treatment, prophylaxis and resistance*.

In the study of these models the following quantities are of interest [19]:

- *Epidemic's peak value and time*; i.e. the maximum number of infectives with respect to time and the time when this occurs.
- *Instantaneous number of total infectives at a time t*:

$$\mathcal{T}_{\text{Infectives}}(t) = \sum_{i=1}^{n} (I_i(t) + I_{s,i}(t) + I_{tr,i}(t) + I_{r,i}(t) + I_{rtr}(t) + I_{sr,i}(t) + I_{srtr,i}(t) + I_{str,i}(t))$$

  where i is an index indicating the location of the species and n is the number of locations. In our study $n = 1$ (no spatial structure) or $n = 5$ (for spatial structure). Note that in the first two models we have just I_i and $I_{s,i}$, so the expression is reduced to $\mathcal{T}_{\text{Infectives}}(t) = \sum_{i=1}^{n} (I_i(t) + I_{s,i}(t))$.
- *Cumulative number of total infectives generated during epidemic within a time t*:

$$\mathcal{T}_{\text{Itot}}(t) = S_0 \; - \; \sum_{i=1}^{n} (S_i(t) + S_{pr,i}(t))$$

  where S_0 is the total number of susceptible individuals over all the locations initially and n and i are as

above. Note the in the first two models we have just the species S_i as we have no prophylaxis.

- *Total number of infectives resistant to drugs at time t* (for models with resistant virus):

$$\mathcal{T}_{\text{Resistants}}(t) = \sum_{i=1}^{n} (I_{r,i}(t) + I_{rtr,i}(t) + I_{sr,i}(t) + I_{srtr,i}(t))$$

Furthermore, we investigate the impact of the population structure on the outcome of the influenza and the effect of treatment and prophylaxis on drug resistance.

5.2 SIIsR models without drug treatments and with single location

The first kind of model is an SIIsR without drug treatments and explicit locations. The species involved are S, I, I_s and R and the interactions are:

(i) contact between S and I described as $S + I \to 2I$, with a contact parameter β_1 (*contact1*);

(ii) contact between S and I_s described as $S + I_s \to I_s + I$, with a contact parameter β_2 (*contact2*);

(iii) the appearance of symptoms in I, described as $I \to I_s$, with rate δ_1 (*symp*);

(iv) recovery from I, with rate γ_1: $I \to R$ (*recovery1*);

(v) recovery from I_s, with rate γ_2: $I_s \to R$ (*recovery2*).

We define the unique location for all the species as $location\_1 : size = 1$. The set $\mathcal{N}$, containing the information about species, is:

$$S : H = 10;\ M_{Max} = 500,\ location = location\_1; \qquad R : H = 10;\ M_{Max} = 500,\ location = location\_1;$$

$$I : H = 10;\ M_{Max} = 500,\ location = location\_1; \qquad I_s : H = 10;\ M_{Max} = 500,\ location = location\_1;$$

the species components' names correspond to the names of the classes. We assume the step size (H) of 10 and a maximum number of individuals (M_{Max}) for all the species equal to the total population.

The functional rates are defined as:

$$f\alpha_{contact1} = \beta_1 \cdot S \cdot I; \qquad f\alpha_{contact2} = \beta_2 \cdot S \cdot I_s; \qquad f\alpha_{symp} = \delta_1 \cdot I; \qquad f\alpha_{recovery1} = \gamma_1 \cdot I; \qquad f\alpha_{recovery2} = \gamma_2 \cdot I_s;$$

where the action types correspond to the interactions listed above and the parameters are as reported in the paper [19]: $\delta_1 = 0.5\, d^{-1}$; $\gamma_1 = 0.5\, d^{-1}$; $\gamma_2 = 0.25\, d^{-1}$; $\beta_1 = 3.5 \cdot 10^{-2}$; $\beta_2 = 5 \cdot 10^{-3}$.

The species components are:

$$S \stackrel{def}{=} (contact1, 1)\downarrow S + (contact2, 1)\downarrow S \qquad R \stackrel{def}{=} (recovery1, 1)\uparrow R + (recovery2, 1)\uparrow R$$

$$I \stackrel{def}{=} (contact1, (1, 2)) \odot I + (contact2, 1)\uparrow I + (recovery1, 1)\downarrow I + (symp, 1)\downarrow I$$

$$I_s \stackrel{def}{=} (contact2, (1, 1)) \odot I_s + (recovery2, 1)\downarrow I_s + (symp, 1)\uparrow I_s$$

and the model component is: $S(s_0) \underset{*}{\bowtie} I(i_0) \underset{*}{\bowtie} I_s(is_0) \underset{*}{\bowtie} R(r_0)$, with the initial values $s_0 = 450$, $i_0 = 10$, $is_0 = 40$, $r_0 = 0$.

Some analysis results are reported in Figs. 4, 5 and 6. Fig. 4 reports the behaviour of all the species obtained from numerical integration of the associated system of ODEs and stochastic simulation (average over 100 runs). The results obtained from the ODEs and stochastic simulation are in full agreement: sus-

ceptibles are infected rapidly and drop to zero very quickly. By time 20 days all the infectives become recovered. Indeed there is not a large variability between the different simulation runs (Fig. 5) and the ODEs are therefore a good approximation for the behaviour of the system. Fig. 6 shows the instantaneous number of infectives and the cumulative number of infective cases up to time t of the epidemic, with the original contact parameters and with smaller values ($\beta_1 = 3.5 \cdot 10^{-3}$, $\beta_2 = 5 \cdot 10^{-4}$). With the original contact parameters the peak of infection happens quickly and the extinction of the disease is rapid. All the susceptibles are infected. With smaller values for the constant parameters, the cumulative number of infectives is smaller and the peak happens later. Furthermore, not all the susceptibles are infected.

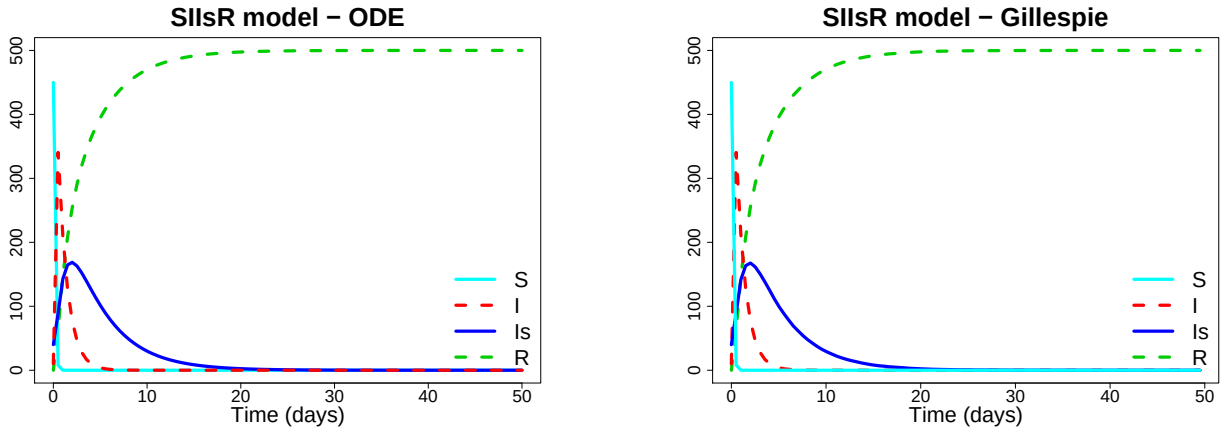

Fig. 4. Simulation results for the $SIIsR$ model without drug treatment and single location: all species (ODE, left), all species (Gillespie's direct method, average of 100 runs, right).

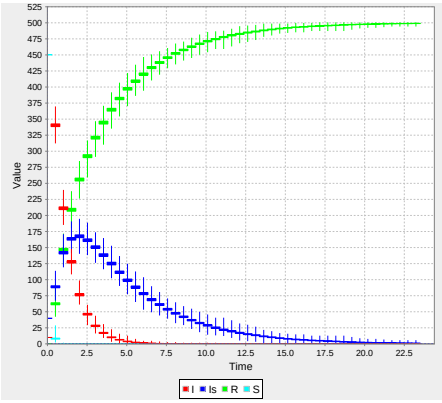

Fig. 5. Simulation results for the $SIIsR$ model without drug treatment and single location: confidence intervals, all species (Gillespie's direct method, average of 100 runs). The thick lines (bars) represent the endpoint of the confidence intervals (with confidence level $1 - \alpha = 0.95$) whereas the thin lines (shadows) illustrates the highest and the lowest values obtained from the set of simulation runs.

Finally, the PRISM model is considered to verify the properties listed in Sect. 4.3, under different assumptions about the contact parameters. The results are reported in Fig. 7. For the original values, the probability of coexistence of all the species is non-zero immediately after the initial time and then drops to zero quickly. The probability of extinction of the disease increases from zero to one. The long-run probabilities are zero and one, respectively, whereas the expected time of extinction of the disease is 12 days. In the case of smaller values for the contact parameters, the probability of coexistence is greater than zero for longer, but then it drops to zero. The probability of extinction tends to one, but more slowly than in the

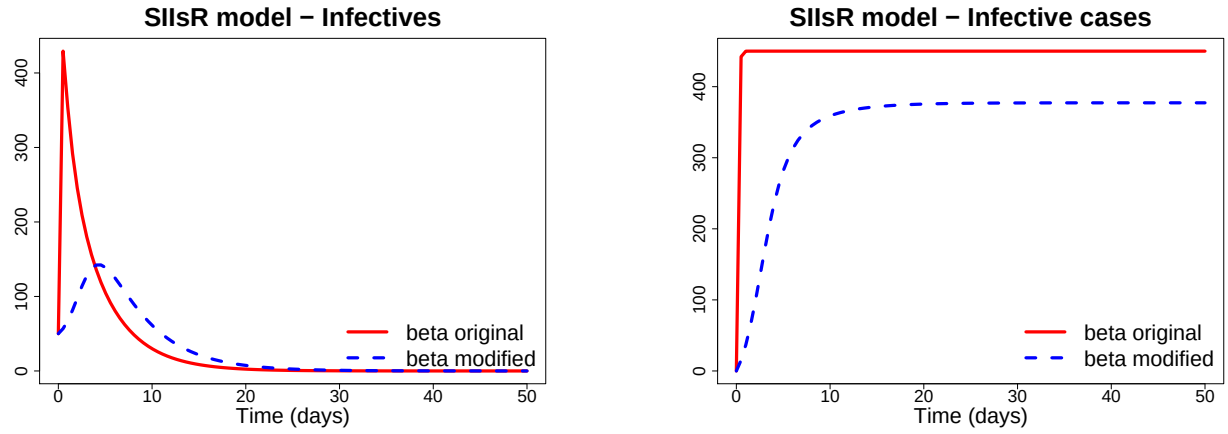

Fig. 6. Simulation results for the $SIIsR$ model without drug treatment and single location: instantaneous number of total infectives (left) and cumulative number of infective cases (right). Standard values for the contact parameters ($\beta_1 = 3.5 \cdot 10^{-2}$, $\beta_2 = 5 \cdot 10^{-3}$) against smaller values ($\beta_1 = 3.5 \cdot 10^{-3}$, $\beta_2 = 5 \cdot 10^{-4}$).

original case. The expected time of extinction is 15 days. These results are in full agreement with the results obtained from the previous analyses and the expected behaviour of this kind of epidemiological systems.

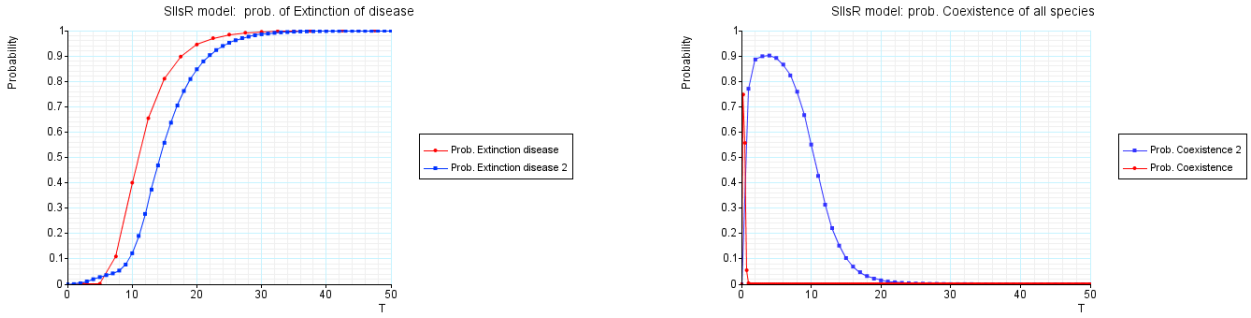

Fig. 7. Analysis obtained from PRISM for the $SIIsR$ model without drugs and single location: probability of extinction of the disease (left) and probability of coexistence of all the species at each time (right), under different assumptions on the contact parameters. The red line shows the results corresponding to the original values and the blue line, the smaller values. These results are based on a CTMC with levels and step size of 10.

5.3 SIIsRs model without drug treatments and with multiple locations

The second considered model is an SIIsR model with 5 patches, initially with 100 individuals each. We analyse models with different spatial structures (*spider-type, loop-type, island-type, necklace type*). All the patches have the same behaviour in terms of the kind of interactions and rates, but the initial situation in terms of disease can be different. Specifically, all the infectives are initially in one single patch i (as depicted in Figure 1). Initially in this patch we have $I_i = 10$, $Is_i = 40$, $S_i = 50$, $R_i = 0$. The other patches initially have 100 susceptible individuals.

The Bio-PEPA model for the SIIsR model with single location in Sect. 5.2 is extended in order to consider a location for each patch, the species for all the locations and the action types for migration between locations. The Bio-PEPA specification of this second model is reported briefly below.

A location is defined for each patch [6] :

[6] We assume size 1 for all the locations as this information is not important here.

$$location\_1 : \ size \ = 1; \ \ location\_2 : \ size \ = 1; \ \ \cdots \ \ location\_5 : \ size \ = 1;$$

As usual, each interaction is associated with an action type and a functional rate. For the parameters, the same interaction in different locations has the same rate, therefore the parameters are as for the previous model, with the addition of the migration rates (μ).

In Bio-PEPA species in different locations are considered different species, so if there are 5 locations we have $5 \cdot 4 = 20$ species components. The class X_i is represented in Bio-PEPA by the species component $X@location_i$. All the classes are listed in the set $\mathcal{N}$ as seen for the model without location. The species components representing the same class in different locations can be defined using the "compact form" presented in [14]: a unique component for the same class in all the possible locations is defined and the interactions that are common to all the locations are represented by a unique action type. If an action type is associated with just one location, the information about the location is added to the prefix term. In the case of susceptibles we have the following definition (the other classes are defined similarly):

$$S \ \stackrel{def}{=} \ (contact1, 1)\downarrow S \ + \ (contact2, 1)\downarrow S \ + \ \sum_{i=1}^{5} \sum_{j=1,\, j\neq i}^{5} (m_{ij,S}[location\_i \rightarrow location\_j], (1,1)) \odot S$$

The action types $contact\,j$, $j = 1, 2$, abstract the action types $contact\,j@location\_i$, with $i = 1, ..., 5$, one for each location, whereas the prefix $m_{(ij),S}[location\_i \rightarrow location\_j]\odot$, $i, j = 1, \cdots, 5$, $i \neq j$, indicates the migration of S from the patch i to the patch j. All the possible migrations between any two patches are listed; when a migration is not possible the value 0 is assigned to the associated migration rate μ, otherwise $\mu = 0.01$. This allows us to easily modify the model in order to consider the different spatial structures.

Finally, the model component is:

$$S\,@location\_1(s1_0) \underset{*}{\bowtie} S\,@location\_2(s2_0) \underset{*}{\bowtie} \cdots S\,@location\_5(s5_0) \underset{*}{\bowtie} I@location\_1(i1_0) \underset{*}{\bowtie} \cdots$$

Figs. 8, 9 and 10 show the results for the island-type structure. Fig. 8 reports deterministic and stochastic simulations for the total number of S, I, Is and R over all the locations; Fig. 9, the confidence intervals for the species over 100 runs of stochastic simulation; Fig. 10 shows the instantaneous number of total infectives and the cumulative number of infective cases. Figs. 11, 12 and 13 report the results for the necklace-type structure. The results for the spider-type structure (not shown) are in agreement with those for the island-type and the results for the loop-type (not shown) are similar to those for the necklace-type.

From Figs. 8 and 11 we can observe that there are some differences between the results obtained from numerical integration of ODEs and Gillespie's average simulation when locations are considered, in contrast to the single location case (Fig. 4). In particular, for stochastic simulation the peak of the disease is less evident than in the basic SIIsR model. This is especially clear for the necklace-type structure; in this case we can observe a large variability between the runs (Fig. 12). Although less pronounced, variability is still evident for the island-type structure (Fig. 9). The presence of spatially separated patches means that the population is fragmented and number of individuals in each location is smaller (initially just 100). Thus we would expect greater variability and less justification for the deterministic fluid approximation offered by ODEs. Thus we focus on stochastic simulation to study the instantaneous number of total infectives and cumulative number of infective cases (see Figs. 10 and 13), for these cases.

Furthermore, note that in the case of subpopulations, susceptibles are infected less rapidly than in a single population (Figs. 8 and 11) and the peak of infectives is lower (Figs. 10 and 13).

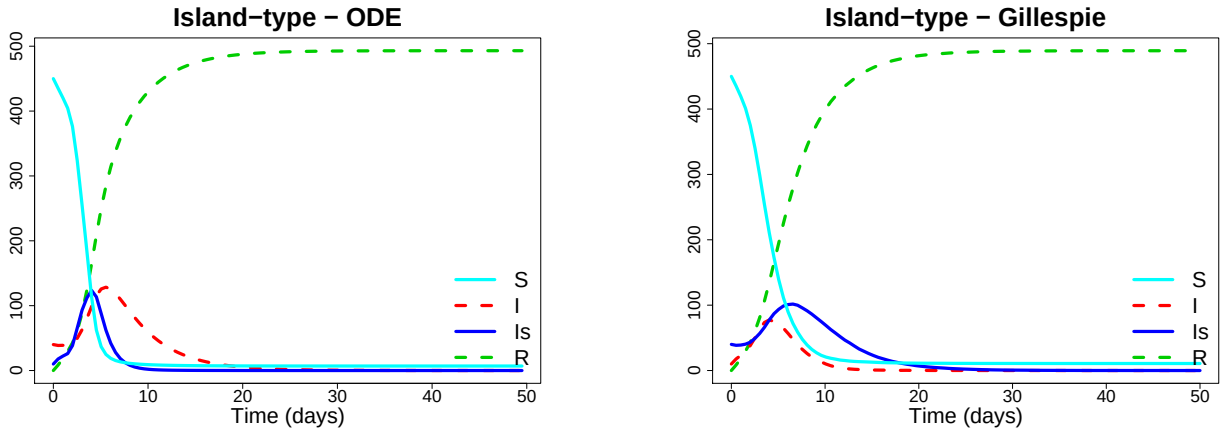

Fig. 8. Simulation results for the $SIIsR$ model with locations (island-type) and without drug treatment: all species (ODE, left), all species (Gillespie's direct method, average of 100 runs, right).

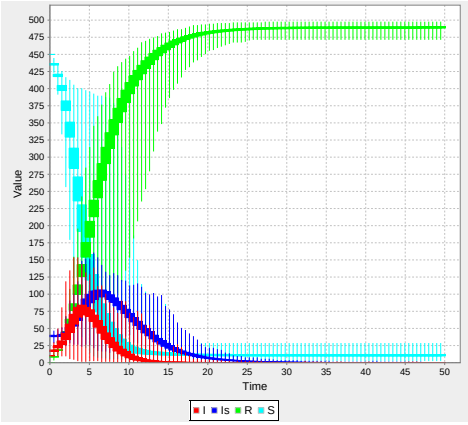

Fig. 9. Simulation results for the $SIIsR$ model with locations (island-type) and without drug treatment: confidence intervals, all species (Gillespie's direct method, average of 100 runs). The thick lines (bars) represent the endpoint of the confidence intervals (with confidence level $1 - \alpha = 0.95$) whereas the thin lines (shadows) illustrates the highest and the lowest values obtained from the set of simulation runs.

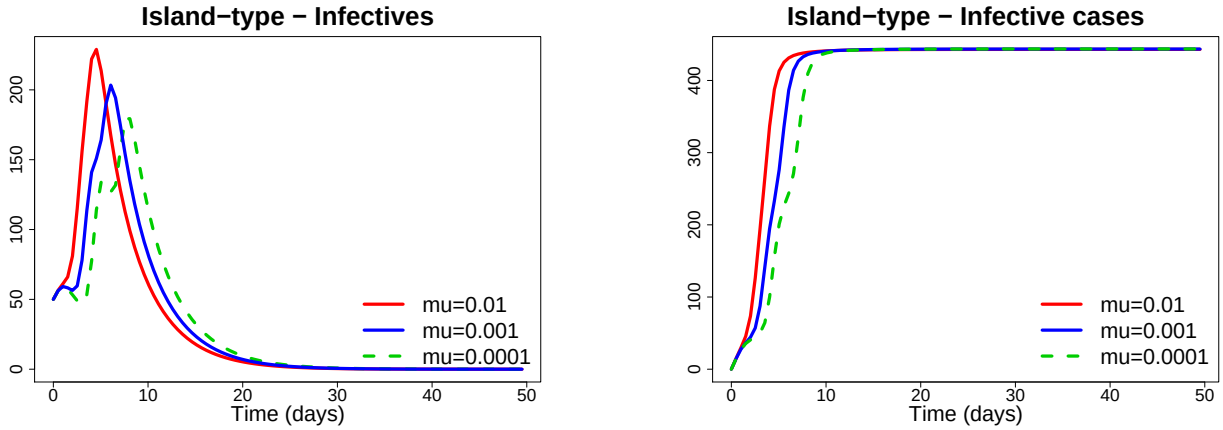

Fig. 10. Simulation results for the $SIIsR$ model with locations (island-type) and without drug treatment: instantaneous number of infectives at a given time point (left) and cumulative number of infectives(right) under different assumptions for the migration rate. Red line is the result for $\mu = 0.01$, blue line for $\mu = 0.001$ and dashed green line for $\mu = 0.0001$.

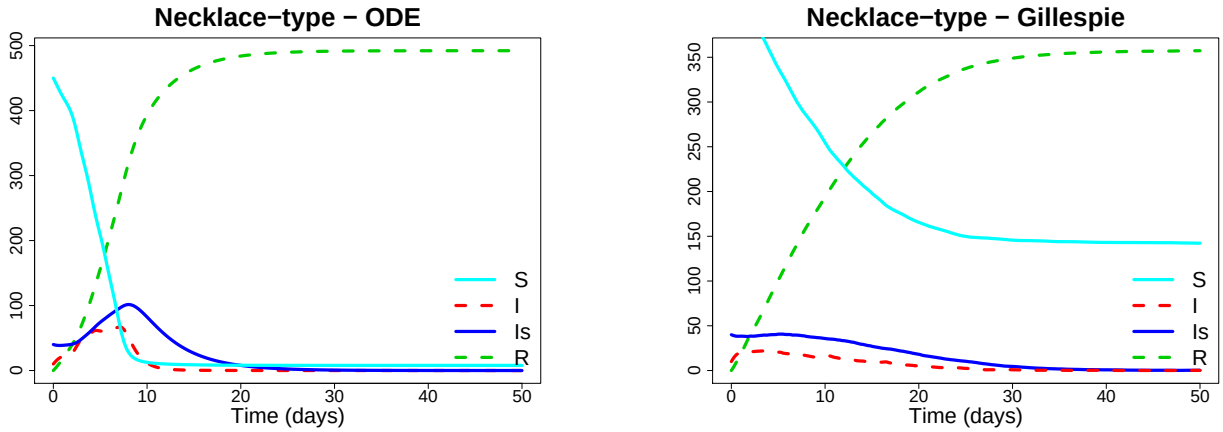

Fig. 11. Simulation results for the $SIIsR$ model with locations (necklace-type) and without drug treatment: all species (ODE, left), all species (Gillespie's direct method, average of 100 runs, right).

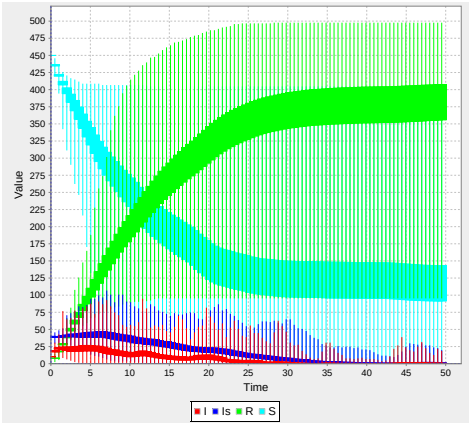

Fig. 12. Simulation results for the $SIIsR$ model with locations (necklace-type) and without drug treatment: confidence intervals (Gillespie's direct method, average of 100 runs). The thick lines (bars) represent the endpoint of the confidence intervals (with confidence level $1-\alpha = 0.95$) whereas the thin lines (shadows) illustrates the highest and the lowest values obtained from the set of simulation runs.

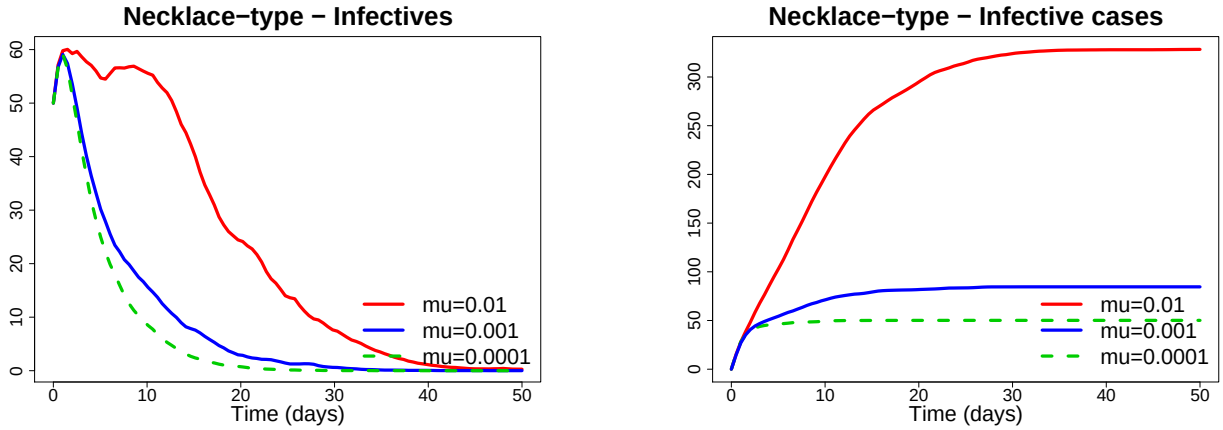

Fig. 13. Simulation results for the $SIIsR$ model with locations (necklace-type) and without drug treatment: instantaneous number of total infectives at a given time point (left) and cumulative number of infective cases (right). Red line is the result for $\mu = 0.01$, blue line for $\mu = 0.001$ and dashed green line for $\mu = 0.0001$. The results show the average over 100 runs of stochastic simulation, Gillespie's direct method.

Moreover, we can observe that when locations are considered, the cumulative number of infectives at the end of epidemic is lower than the case without locations; not all susceptibles are infected. In the case of the island-type structure we have about 7 susceptibles that are not infected (Fig. 10). For the necklace-type structure we have even more uninfected susceptibles during the epidemic (about 110 cases for the original migration rates (see Fig 13)). These results are as expected: the epidemic is less explosive in a fragmented population than in a homogeneous one.

Finally, in the case of island-type structure, the epidemic peak decreases and occurs later with decreasing migration rates, but the total number of infective cases at the end of epidemic does not depend on the migration rates (Fig. 10). This is not true for the necklace-type structure: the cumulative number of total infectives at the end of epidemic is influenced by the migration rates: the lower the migration rate, the lower the cumulative number of infectives.

These observations are in agreement with the results shown in the literature [19].

5.4 SIIsR with locations, treatment and drug resistance

The models presented in the previous section can be extended in order to consider *treatment, resistance to drugs* and *prophylaxis*. We consider all the 11 species defined in Fig. 3. The possible interactions are:

- As before, contact between a susceptibles and an infective results in a new infective but now we distinguish different types. Globally, there are 16 possible kinds of contact interactions. If the susceptible S contacts an infective without resistance the result is a simple infective I, if S contacts an infective with resistance the result is I_r. For a susceptible with prophylaxis, the contact with an infective without resistance is I_{tr}, whereas if the contact is with an infective with drug resistance the result is I_{rtr}. Similar interactions are defined for all the other cases of infectives. These interactions are described by appropriate transmission rates. The transmission rates for the treated infectives are lower than the usual ones (we assume that they are 30 per cent of the usual ones).
- *Development of symptoms*. Asymptomatic infectives can become symptomatic after some time.
- *Treatment.* Treatment is for all the symptomatic individuals. The treatment does not work immediately but with a certain delay ($0.7\ d^{-1}$).
- *Recovery*. All the infectives can recover.
- *Drug resistance*. Infectives that are treated and non-resistant to drugs can become resistant to them.

In addition to these interactions, prophylaxis is added at time 1 day after the beginning of the study [7] and involves just the patches with disease. We consider that only a part of the population is subjected to prophylaxis (one third of the susceptibles and asymptomatic infectives). Furthermore, the transmission rates for individuals with prophylaxis is just 30 per cent of the usual one.

In this model we assume an island-type structure with 5 patches. Each patch has 100 individuals and initially all the infectives are in the same patch. The initial situation for this patch is 50 susceptibles, 50 total infectives (asymptomatic and symptomatic) and all the other species are zero.

The Bio-PEPA model corresponding to this case it not reported. The translation is similar to the simpler models, but more species and action types are added to describe the new species and interactions. The main

[7] We assume that the infection has started some days before the start of the simulation and therefore prophylaxis is added, realistically, some days after the spread of the disease.

difference is the addition of an event representing the prophylaxis at time 1 day of the simulation time:

$$(prophylaxis,\ t = 1\ day,\ \ S@location\_i \leftarrow S@location\_i \cdot 2/3;\ S_{pr}@location\_i \leftarrow S@location\_i \cdot 1/3;$$
$$I@location\_i \leftarrow I@location\_i \cdot 2/3;\ I_{tr}@location\_i \leftarrow I@location\_i \cdot 1/3;$$
$$I_r@location\_i \leftarrow I_r@location\_i \cdot 2/3; I_{rtr}@location\_i \leftarrow I_r@location\_i \cdot 1/3,\ i = 1,2,3,4,5,\ \ 0)$$

The event *prophylaxis* is immediate and the effect is that one third of susceptibles and asymptomatic infectives become treated susceptibles and treated asymptomatic infectives, respectively. This is obtained by resetting the number of individuals of the species components concerned.

The analysis results for this model are shown in Figs. 14 and 15. Again, the results obtained from Gillespie's simulation are somewhat different from those obtained from the ODEs. In particular, a significant variability between the various simulation runs is present (not shown) and can be explained as for model 2. Furthermore, a larger number of susceptibles resist infection compared with the scenarios without treatments and therefore the total number of infectives is less. The infection peak is lower than the peak for the model with the same structure and without treatment. Finally, there are some resistant cases and there is quite quickly a peak of drug resistance (see Fig. 15).

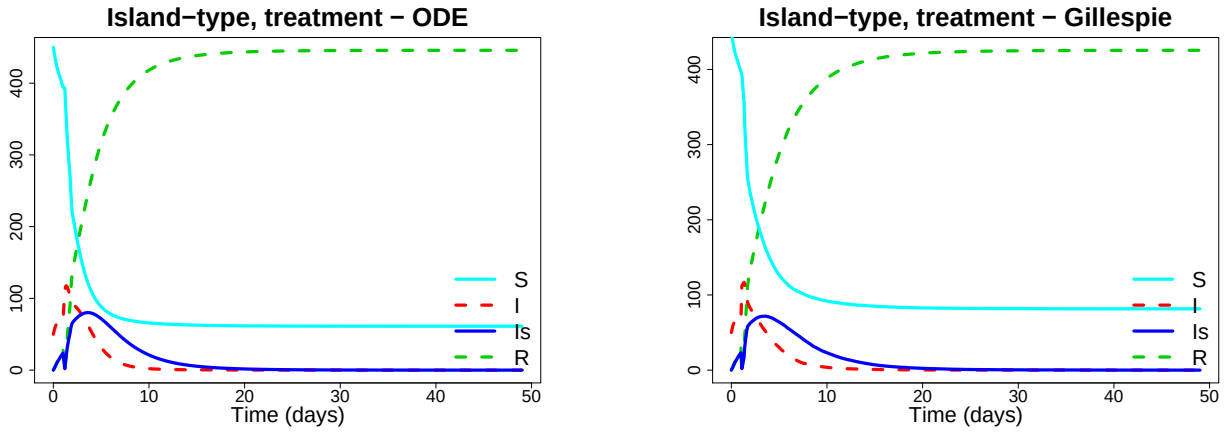

Fig. 14. Simulation results for the $SIIsR$ model with treatment/prophylaxis, drug resistance and locations (island-type): all species (ODE, left) and all species (Gillespie, average over 100 runs, right).

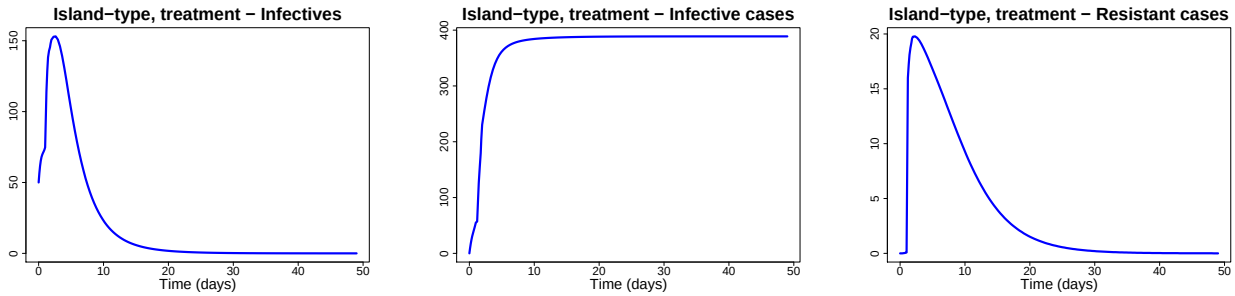

Fig. 15. Simulation results for the $SIIsR$ model with treatment/prophylaxis, drug resistance and locations (island-type): instantaneous number of total infectives at a given time point (left), cumulative number of total infectives (middle) and total resistant cases (right).

6 Conclusions

In this paper we proposed a variant of the process algebra Bio-PEPA for the modelling and analysis of epidemiological models. Bio-PEPA has been shown to be useful in this context: it offers a high level of abstraction and allows us to represent easily features of epidemiological systems, such as complex dynamic laws, abstract spatial structures and changes to the system due to some trigger conditions. These features are represented in Bio-PEPA using functional rates, locations and events, respectively.

Note that there is a vast literature of models for epidemiology and most of them are defined directly in terms of systems of ODEs. However, in recent years there has been some interest in the use of process algebras for the modelling and analysis of epidemiological models [33,28,27,4,7]. In [33,28,27] the WSCCS (Weighted Synchronous Calculus of Communicating Systems) process algebra has been used. In [28,27] McCaig *et al*. define a systematic mapping from WSCCS to the ODEs generally associated with epidemiological models. In [4] Benkirane et al. developed some PEPA models of disease spread (SIR models with direct and indirect transmission) and proposed a new mapping from PEPA to ODEs, able to capture some properties of epidemiogical systems. In [7] Bradley *et al*. applied PEPA to the modelling of Internet worm attacks using the SIR abstraction; analysis was based on the mapping from PEPA to ODEs. In all these cases the models proposed were simple, without spatial structure, and stochastic simulation was used only in [4].

In this work we constructed some models describing the spread of influenza under different assumptions about spatial structure and treatment and we studied them using the various analysis techniques supported by Bio-PEPA. The use of different kinds of analysis can foster a better understanding of the behaviour of the system, and help to discover errors due to the use of a particular solver/simulator [9]. Furthermore the modeller can select the approach that is most appropriate for specific model under study. The first model considered assumed a constant population without spatial structure and treatment. The numerical integration of the associated system of ODEs and the stochastic simulation (average of 100 runs) showed the same behaviour. The system was characterized by a small variability and therefore the ODE solution is a good approximation of the behaviour of the system. For this model, a map to PRISM was also considered in order to formally verify some properties. The results of all the analyses were in full agreement. The other models were characterized by some spatial structure and for them the use of stochastic simulation was essential in order to represent the behaviour of the system. Indeed the subpopulations were characterized by a small number of individuals and the variability in the system was significantly increased. In this case it is not clear that the deterministic fluid approximation offered by an ODE model is appropriate.

Future work will concern the study of other epidemiological models in Bio-PEPA, the application of PRISM to the models with spatial structure and the application of further analysis techniques supported by the Bio-PEPA tools, such as sensitivity analysis.

Acknowledgement

This work was completed while Ciocchetta was a research fellow at the University of Edinburgh, supported by the EPSRC grant EP/c54370x/01. Hillston is supported by the EPSRC ARF EP/c543696/01.

References

[1] Anderson, R., "Population Dynamics of Infectious Diseases: Theory and Applications," Chapman and Hall, London-New York, 1982.

[2] Anderson, R. and R. M. May, "Infectious Diseases of Humans," Oxford University Press, 1991.

[3] Aziz, A., K. Sanwal, V. Singhal and R. Brayton, *Verifying continuous time markov chains*, Proc. 8th International Conference on Computer Aided Verification (CAV'96), LNCS **1102** (1996), pp. 269–276.

[4] Benkirane, S., J. Hillston, C. McCaig, R. Norman and C. Shankland, *Improved Continuous Approximation of PEPA Models through Epidemiological Examples*, ENTCS **229 (1)** (2009), pp. 59–74.

[5] Bio-PEPA Home Page, http://www.biopepa.org/.

[6] Blower, S. M., A. R. Mclean, T. C. Porco, P. M. Small, P. C. Hopewell, M. A. Sanchez and et al., *The intrinsic transmission dynamics of tuberculosis epidemics*, Nature Medicine **1** (1995), pp. 815 – 821.

[7] Bradley, J., S. Gilmore and J. Hillston, *Analysing distributed Internet worm attacks using continuous state-space approximation of process algebra models*, J. Comput. Syst. Sci. **74** (2008), pp. 1013–1032.

[8] Brauer, F. and C. Castillo-Chvez, "Mathematical Models in Population Biology and Epidemiology," Springer, 2001.

[9] Calder, M., A. Duguid, S. Gilmore and J. Hillston, *Stronger computational modelling of signalling pathways using both continuous and discrete-state methods*, in: *Proceedins of Computational Methods in Systems Bioogy (CMSB'06)*, Lecture Notes in Computer Science **4210** (2006), pp. 63–77.

[10] Calder, M., S. Gilmore and J. Hillston, *Modelling the Influence of RKIP on the ERK Signalling Pathway Using the Stochastic Process Algebra PEPA*, Transactions on Computational Systems Biology **7** (2006), pp. 1–23.

[11] Ciocchetta, F., *Bio-PEPA with events*, T. Comp. Sys. Biology **XI** (2009), pp. 45–68.

[12] Ciocchetta, F., A. Degasperi, J. Hillston and M. Calder, *Some Investigations Concerning the CTMC and the ODE Model Derived From Bio-PEPA*, ENTCS **229 (1)** (2009), pp. 145–163.

[13] Ciocchetta, F., A. Duguid, S. Gilmore, M. Guerriero and J. Hillston, *The Bio-PEPA Tool Suite*, in: *Proceedings of the 6th International Conference on Quantitative Evaluation of SysTems (QEST 2009)*, Budapest, Hungary, 2009, to appear.

[14] Ciocchetta, F. and M. Guerriero, *Modelling Biological Compartments in Bio-PEPA*, ENTCS **227** (2009), pp. 77–95.

[15] Ciocchetta, F. and J. Hillston, *Bio-PEPA: a framework for the modelling and analysis of biological systems*, Technical report, School of Informatics University of Edinburgh Technical Report EDI-INF-RR-1231 (2008).

[16] Ciocchetta, F. and J. Hillston, *Bio-PEPA: a Framework for the Modelling and Analysis of Biochemical Networks*, Theoretical Computer Science **410** (2009), pp. 45–68.

[17] Curti, M., P. Degano, C. Priami and C. Baldari, *Modelling biochemical pathways through enhanced π-calculus*, Theoretical Computer Science **325** (2004), pp. 111–140.
URL http://dx.doi.org/10.1016/j.tcs.2004.03.066

[18] Daley, D. J. and J. Gani, "Epidemic Modeling and Introduction," Cambridge University Press, 2005.

[19] Debarre, F., S. Bohnoeffer and R. Regoes, *The effect of population structure on the emergence of drug resistence during influenza pandemics*, Ecology **77** (2007), pp. 893–906.

[20] Dizzy Edinburgh version, http://homepages.inf.ed.ac.uk/stg/software/Dizzy/.

[21] Gillespie, D., *Exact stochastic simulation of coupled chemical reactions*, J Phys Chem **81** (1977), pp. 2340–2361.

[22] Hess, G., *Diseases in Metapopulation Models: Implication for conservation*, Ecology **77** (1996), pp. 1617–1632.

[23] Hethcote, H. W., *The mathematics of infectious diseases*, Society for Industrial and Applied Mathematics **42** (2000), pp. 599 – 653.

[24] Hillston, J., "A Compositional Approach to Performance Modelling," Cambridge University Press, 1996.

[25] Hofbauer, J. and K. Sigmund, "The Theory of Evolution and Dynamical Systems," Cambridge University press, 1998.

[26] Kwiatkowska, M., G. Norman and D. Parker, *Prism: Probabilistic model checking for performance and reliability analysis*, ACM SIGMETRICS Performance Evaluation Review (2009).

[27] McCaig, C., "Studies in the use of Process Algebra as an Analytical Tool for Biological Systems," Ph.D. thesis, University of Stirling (2008).

[28] McCaig, C., R. Norman and C. Shankland, *Process Algebra Models of Population Dynamics*, in: *Proc. of Algebraic Biology, 3rd International Conference, AB 2008*, LNCS **5147**, 2008, pp. 139–155.

[29] Priami, C. and P. Quaglia, *Beta binders for biological interactions*, in: *Proceedings of Computational Methods in Systems Biology (CMSB'04)*, LNCS **3082**, 2005, pp. 20–33.
URL http://springerlink.metapress.com/openurl.asp?genre=article\&issn=0302-9743\&volume=3082\&spage=20

[30] Priami, C., A. Regev, W. Silverman and E. Shapiro, *Application of a stochastic name-passing calculus to representation and simulation of molecular processes*, Information Processing Letters **80** (2001), pp. 25–31.
URL http://dx.doi.org/10.1016/S0020-0190(01)00214-9

[31] PRISM Home Page, *http://www.prismmodelchecker.org*.

[32] Regev, A., E. Panina, W. Silverman, L. Cardelli and E. Shapiro, *BioAmbients: an Abstraction for Biological Compartments*, Theoretical Computer Science **325** (2004), pp. 141–167.
URL http://lucacardelli.name/Papers/BioAmbients%20An%20Abstraction%20for%20Biological%20Compartments.pdf

[33] Sumpter, D., "From Bee to Society: an agent based investigation of honeybee colonies," Ph.D. thesis, UMIST (2000).

[34] The Bio-PEPA Workbench, http://www.dcs.ed.ac.uk/home/stg/software/biopepa/about.html.